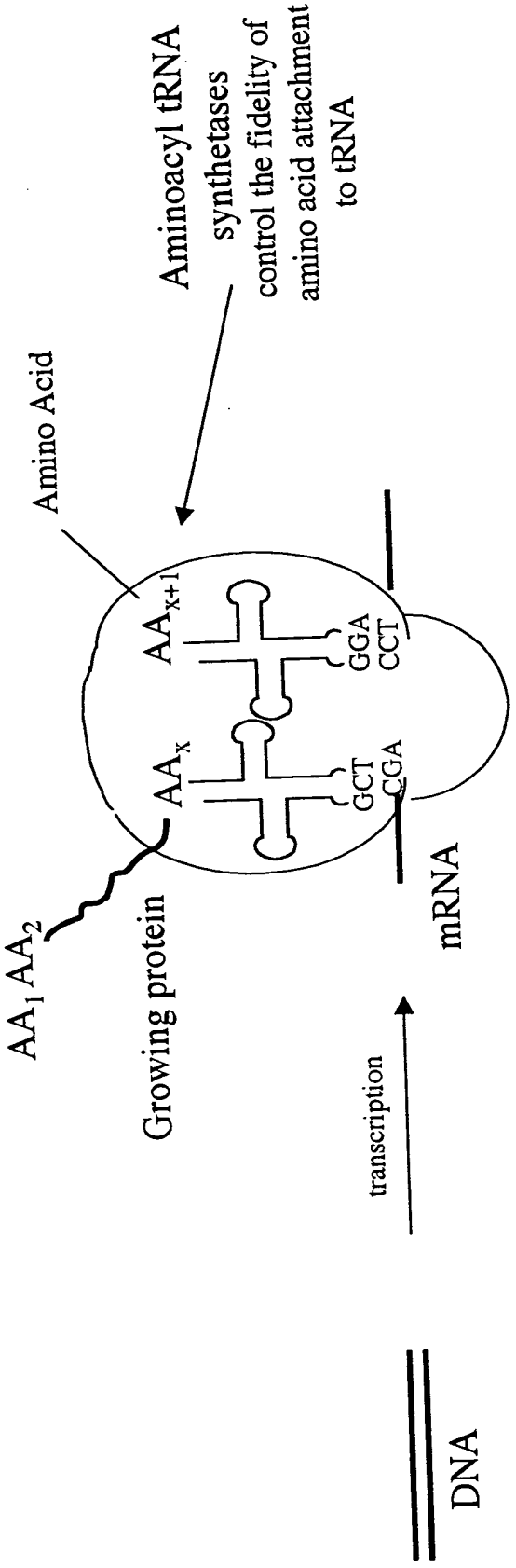
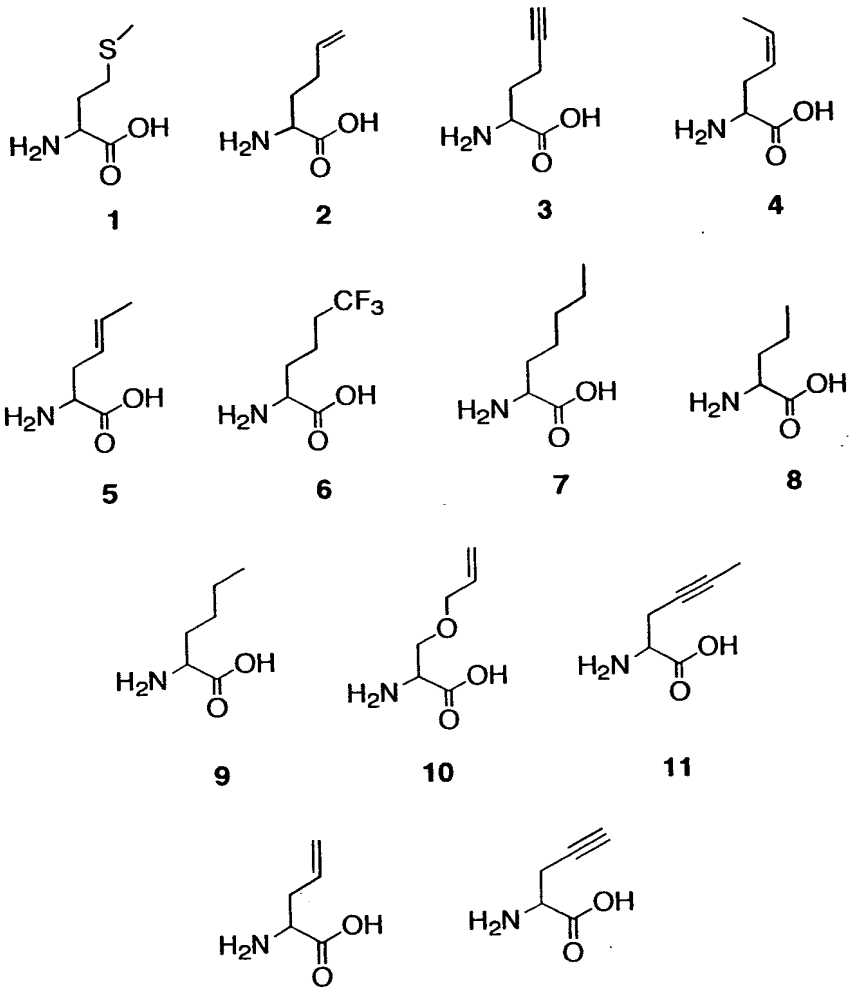

Aminoacyl tRNA synthetases control the fidelity of amino acid attachment to tRNA
Amino Acid
AA_{x+1}
AA_x
GGA
CCT
GCT
CGA
AA_1 AA_2
Growing protein
mRNA
transcription
DNA

FIG. 1



1. **Methionine**
2. **Homoallylglycine**
3. **Homopropargylglycine**
4. **Cis-crotylglycine**
5. **Trans-crotylglycine**
6. **6,6,6-trifluoro-2-amino hexanoic acid**
7. **2-amino heptanoic acid**
8. **Norvaline**
9. **Norleucine**
10. **o-allylserine**
11. **2-butynylglycine**
12. **Allylglycine**
13. **Propargylglycine**

FIG. 2

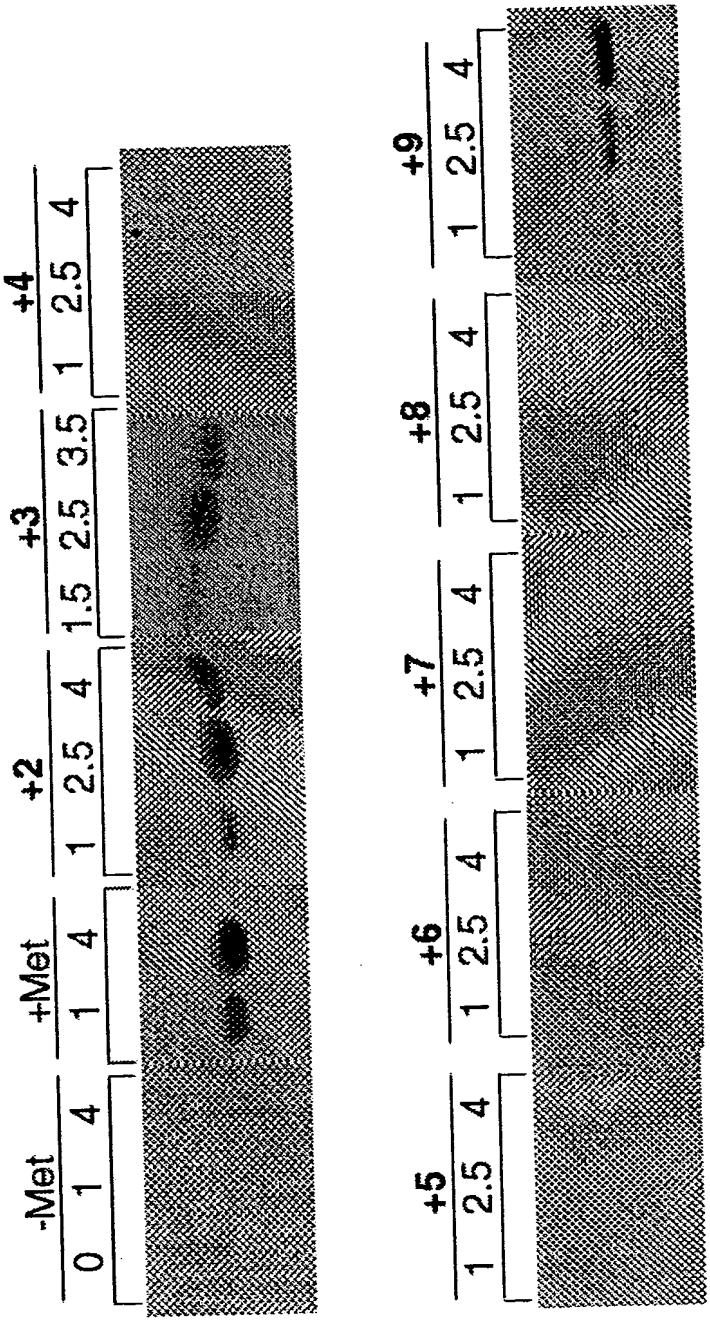

-Met
0 1 4
+Met
1 4
+2
1 2.5 4
+3
1.5 2.5 3.5
+4
1 2.5 4
+5
1 2.5 4
+6
1 2.5 4
+7
1 2.5 4
+8
1 2.5 4
+9
1 2.5 4

FIG. 3

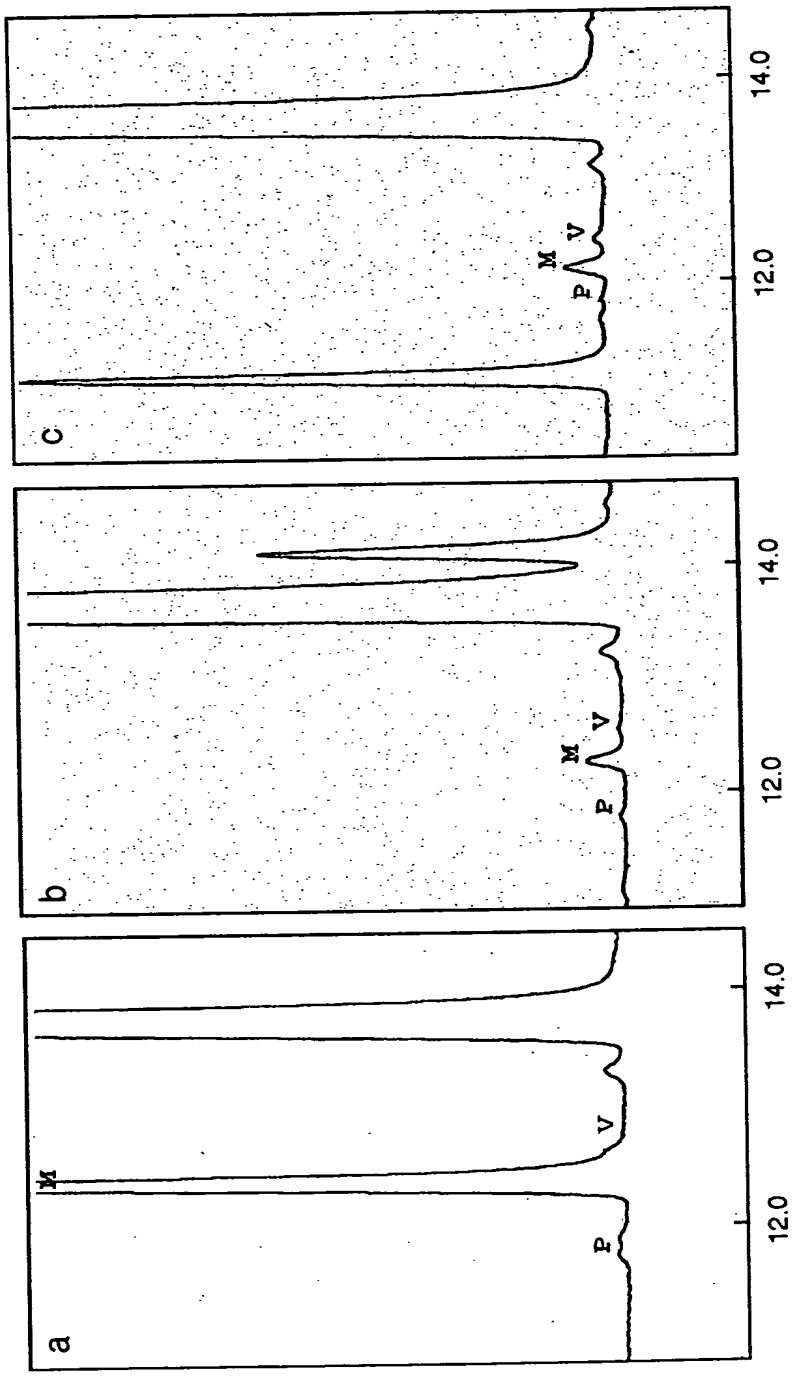


FIG. 4

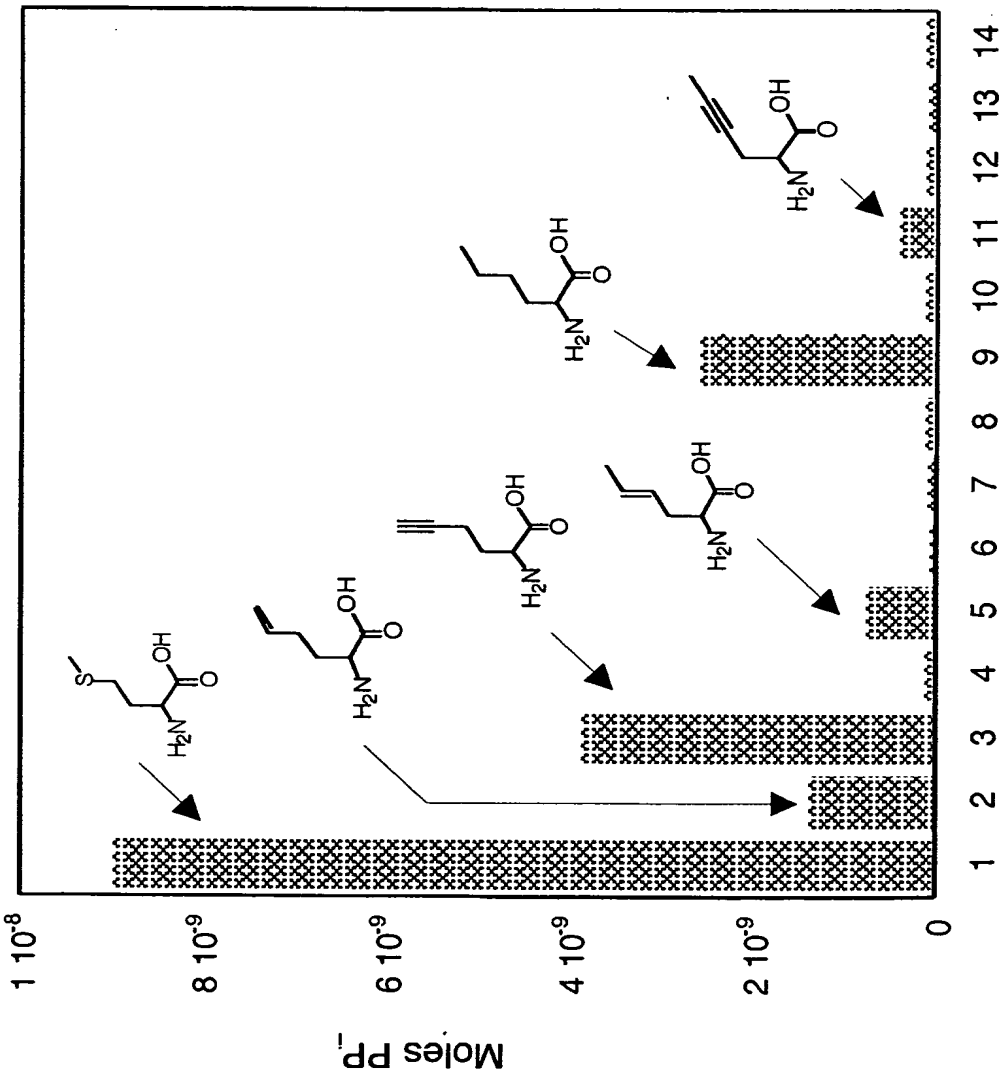

Moles PPi
1 10-8
8 10-9
6 10-9
4 10-9
2 10-9
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Analogue

FIG. 5

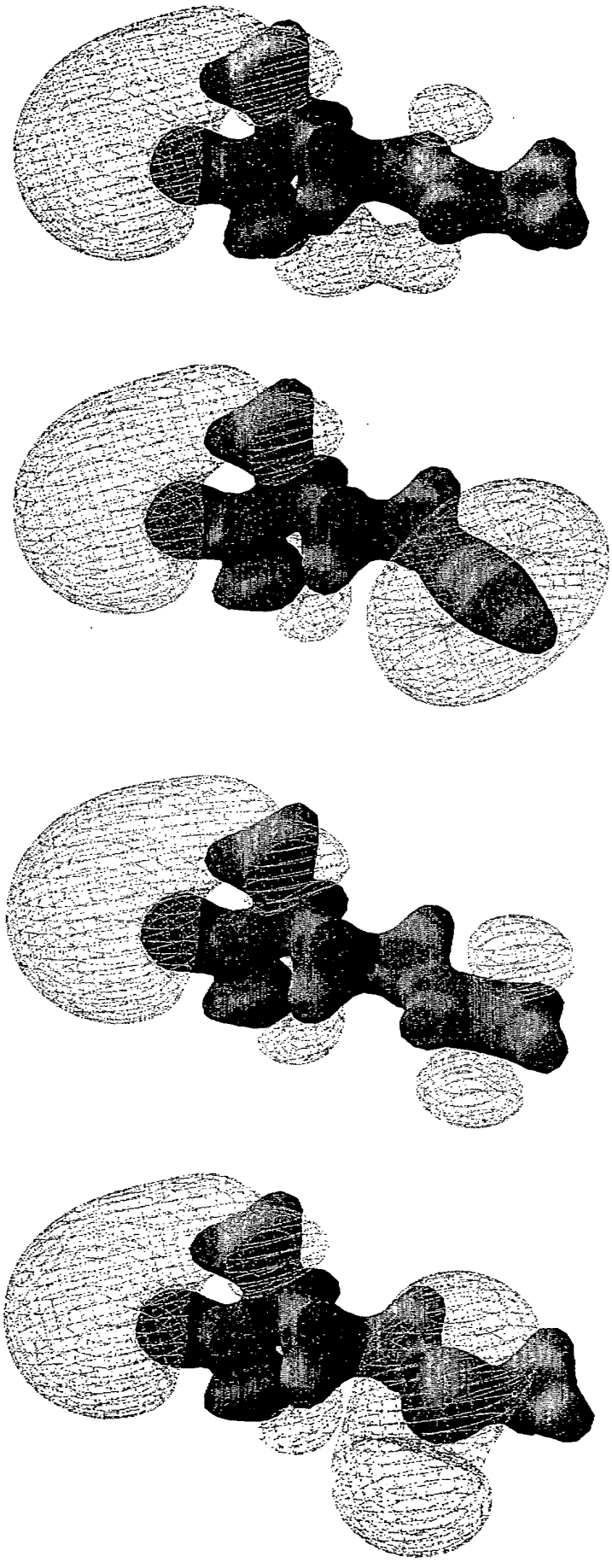


FIG. 6

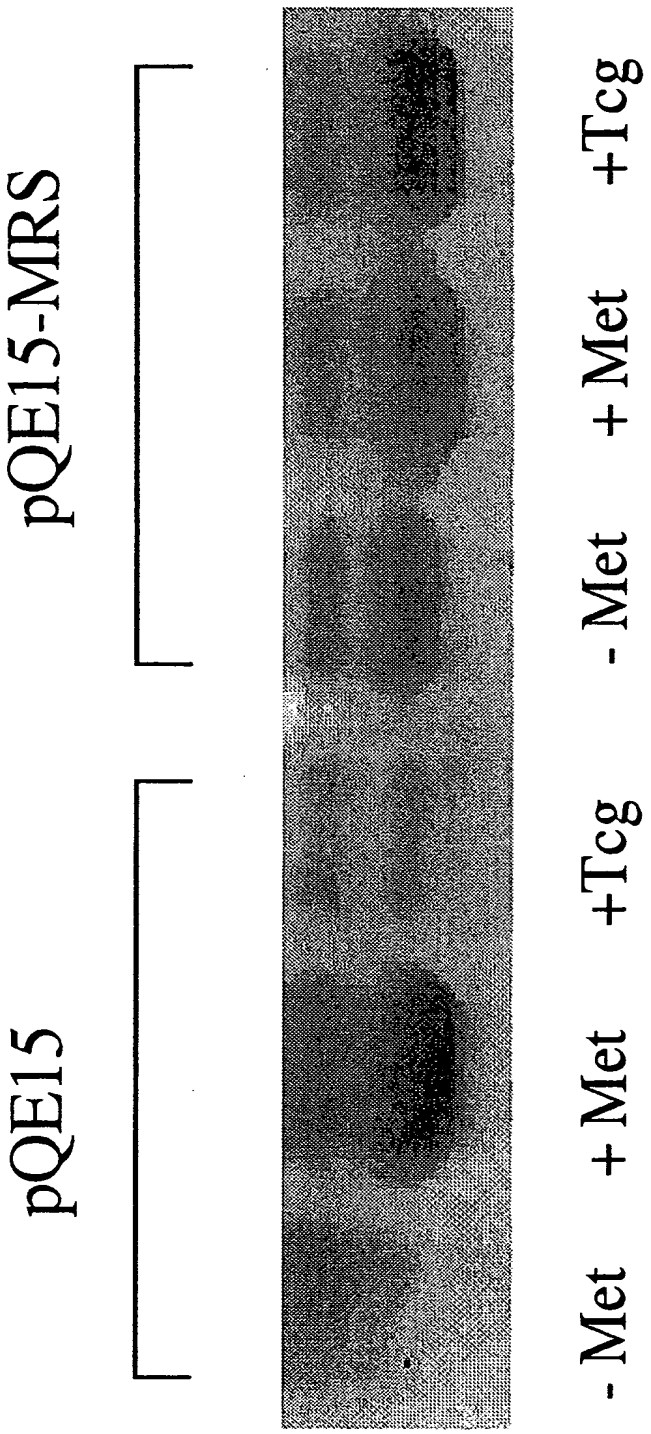


FIG. 7

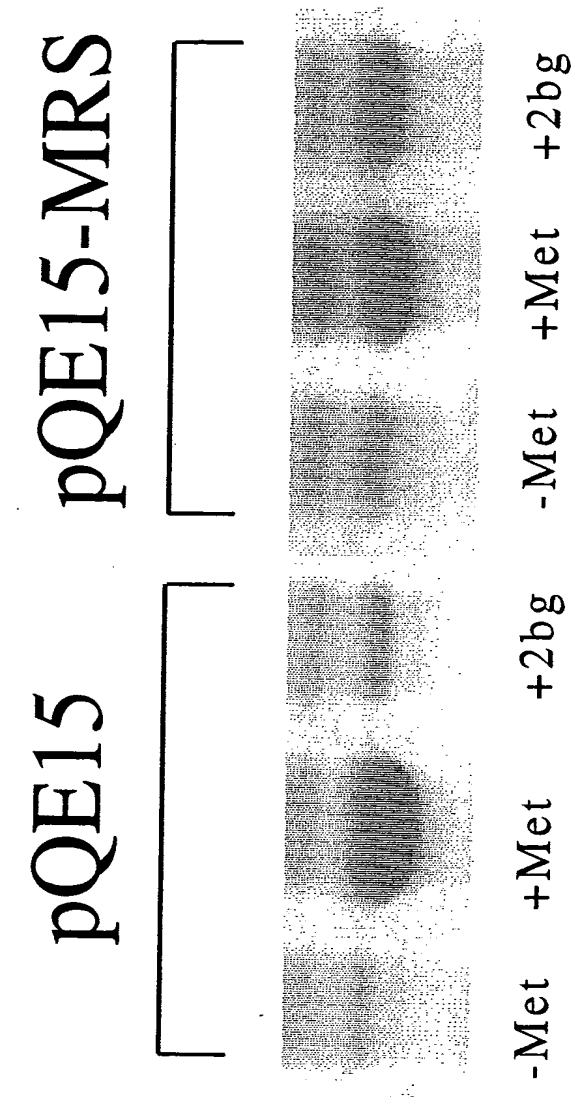


FIG. 8

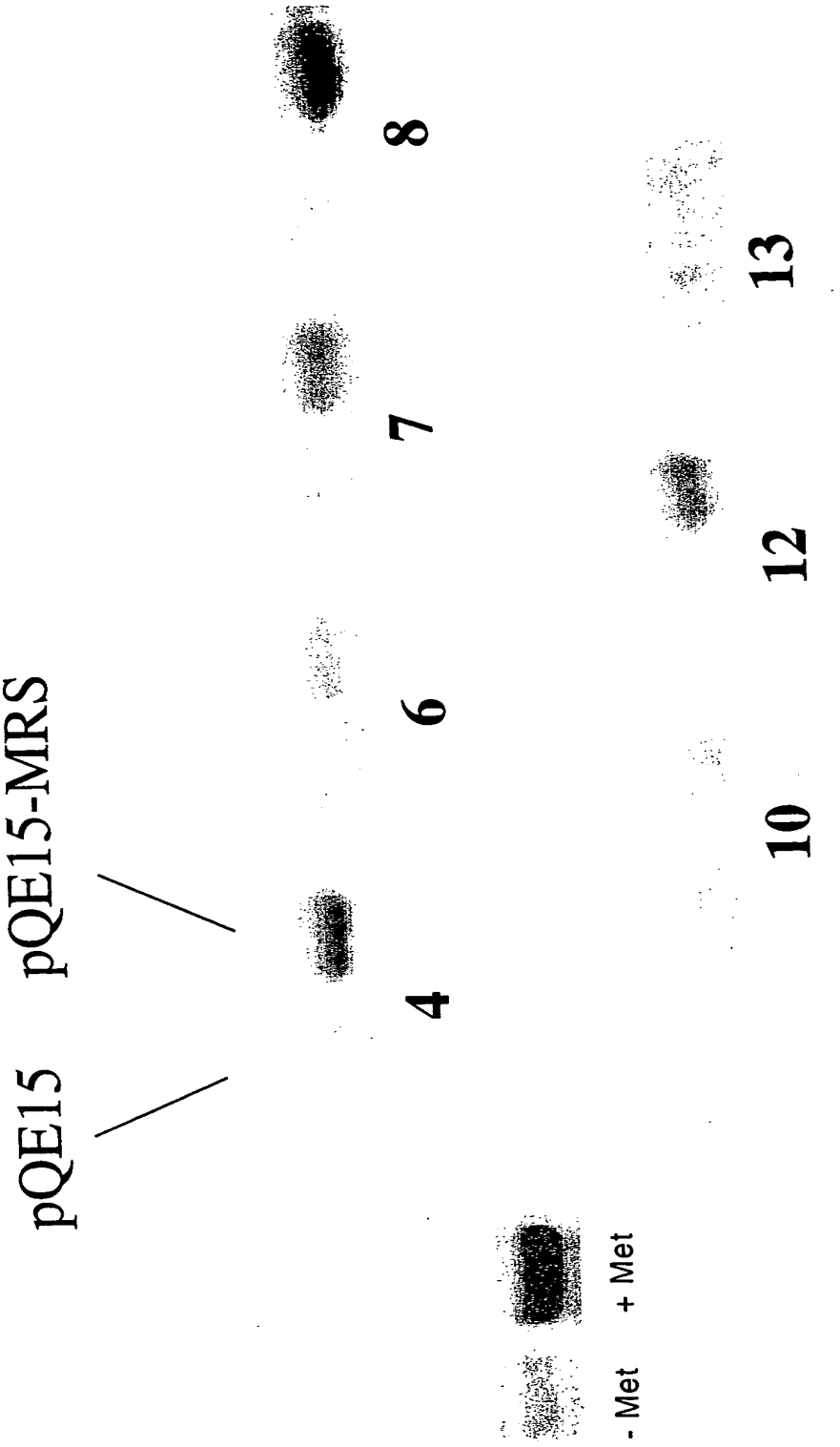


FIG. 9

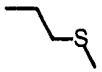
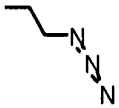
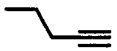
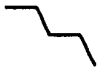
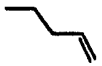
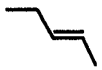
Analogue (Side chain shown)		k_{cat}/K_m ($s^{-1}\,\mu M^{-1}$)	Relative Value	Incorporated by conventional host?
Met		5.47×10^{-1}	1	Y
Aha		1.4×10^{-3}	1/390	Y
Hpg		1.08×10^{-3}	1 / 500	Y
Norl		5.22×10^{-4}	1 / 1050	Y
Hag		2.96×10^{-4}	1 / 1850	Y
Tcg		1.16×10^{-4}	1 / 4700	N
2bg		3.9×10^{-5}	1/13825	N
Norv		1.2×10^{-5}	1/46100	N
Ccg		3.2×10^{-6}	1/171000	N
Ag		1.2×10^{-6}	1/456000	N

FIG. 10

APPLICANT(S): Tirrell and Kiick
TITLE: OVEREXPRESSION OF AMINOACYL-tRNA SYNTHETASES FOR EFFICIENT PRODUCTION OF ENGINEERED PROTEINS CONTAINING AMINO ACID ANALOGUES
Application No.: Unassigned Filing Date: July 1, 2003
Docket No.: CIT1460-2

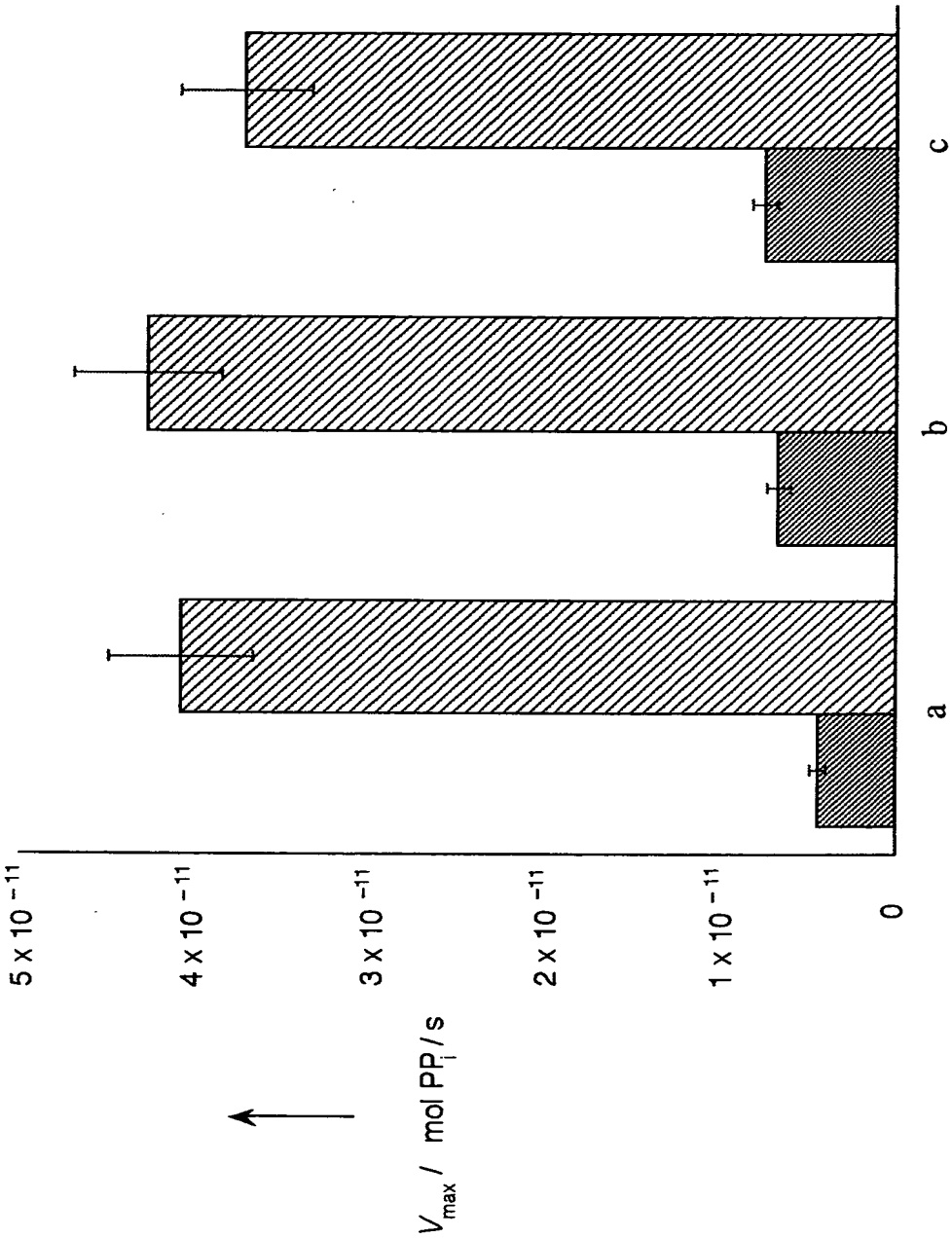


FIG. 11

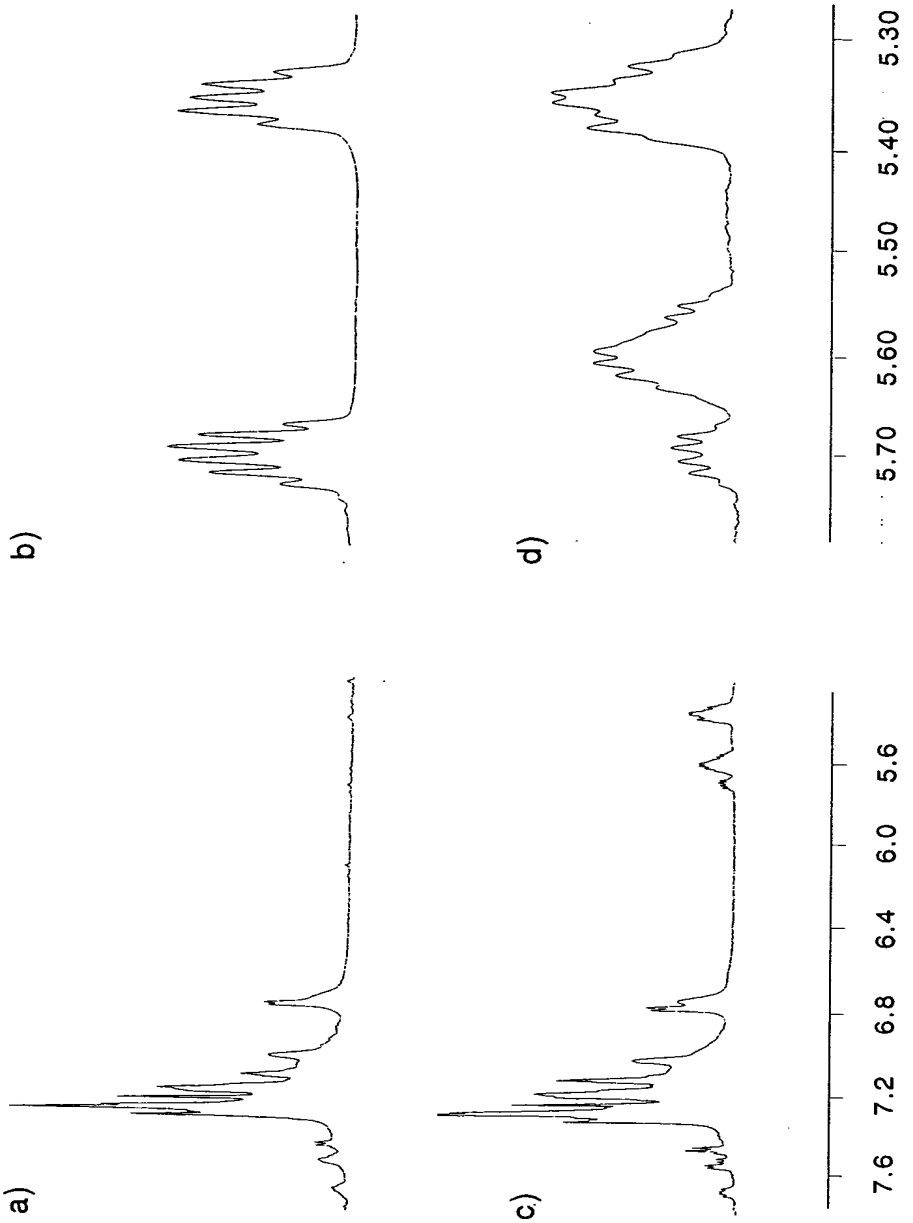


FIG. 12

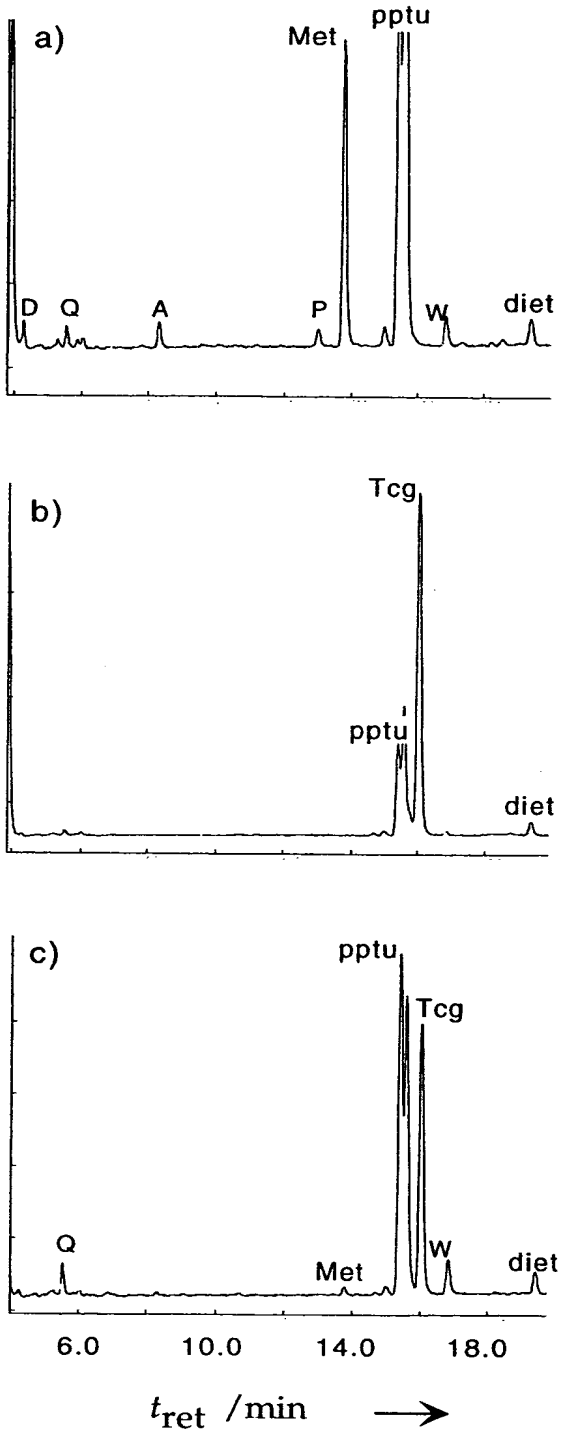

a)
Met
pptu
D
Q
A
P
W
diet
b)
Tcg
pptu
diet
c)
pptu
Tcg
Q
Met
W
diet
6.0
10.0
14.0
18.0
t_{ret} /min

FIG. 13

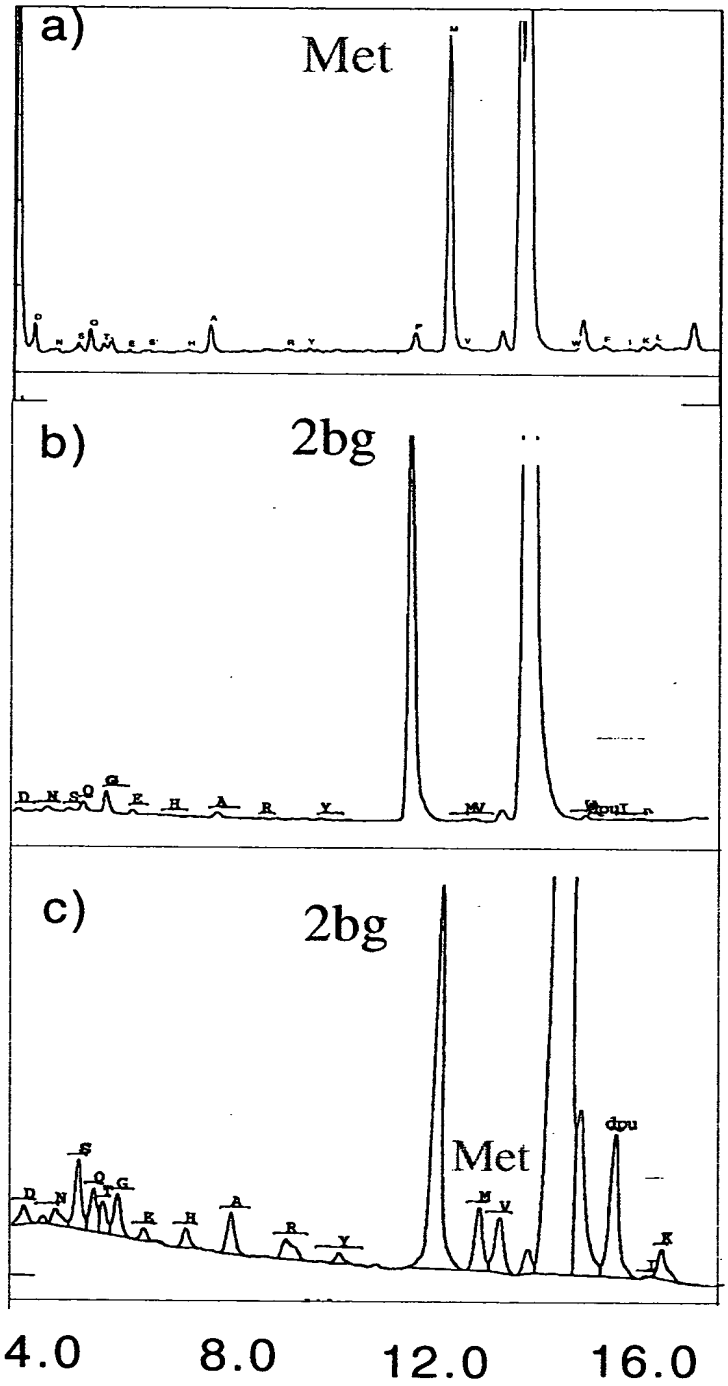


FIG. 14

Analogue	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}\mu M^{-1}$)	Protein Yield, mg/L
1	24.3 ± 2	13.3 ± 0.2	$5.47x10^{-1}$	35
3	2415 ± 170	2.60 ± 0.3	$1.08x10^{-3}$	35
9	4120 ± 900	2.15 ± 0.6	$5.22x10^{-4}$	20
2	4555 ± 200	1.35 ± 0.1	$2.96x10^{-4}$	10
5	15,675 ± 250	1.82 ± 0.6	$1.16x10^{-4}$	0

FIG. 15

APPLICANT(S): Tirrell and Kirck
TITLE: OVEREXPRESSION OF AMINOACYL-tRNA SYNTHETASES FOR EFFICIENT PRODUCTION OF ENGINEERED PROTEINS CONTAINING AMINO ACID ANALOGUES
Application No.: Unassigned Filing Date: July 1, 2003
Docket No.: CIT1460-2

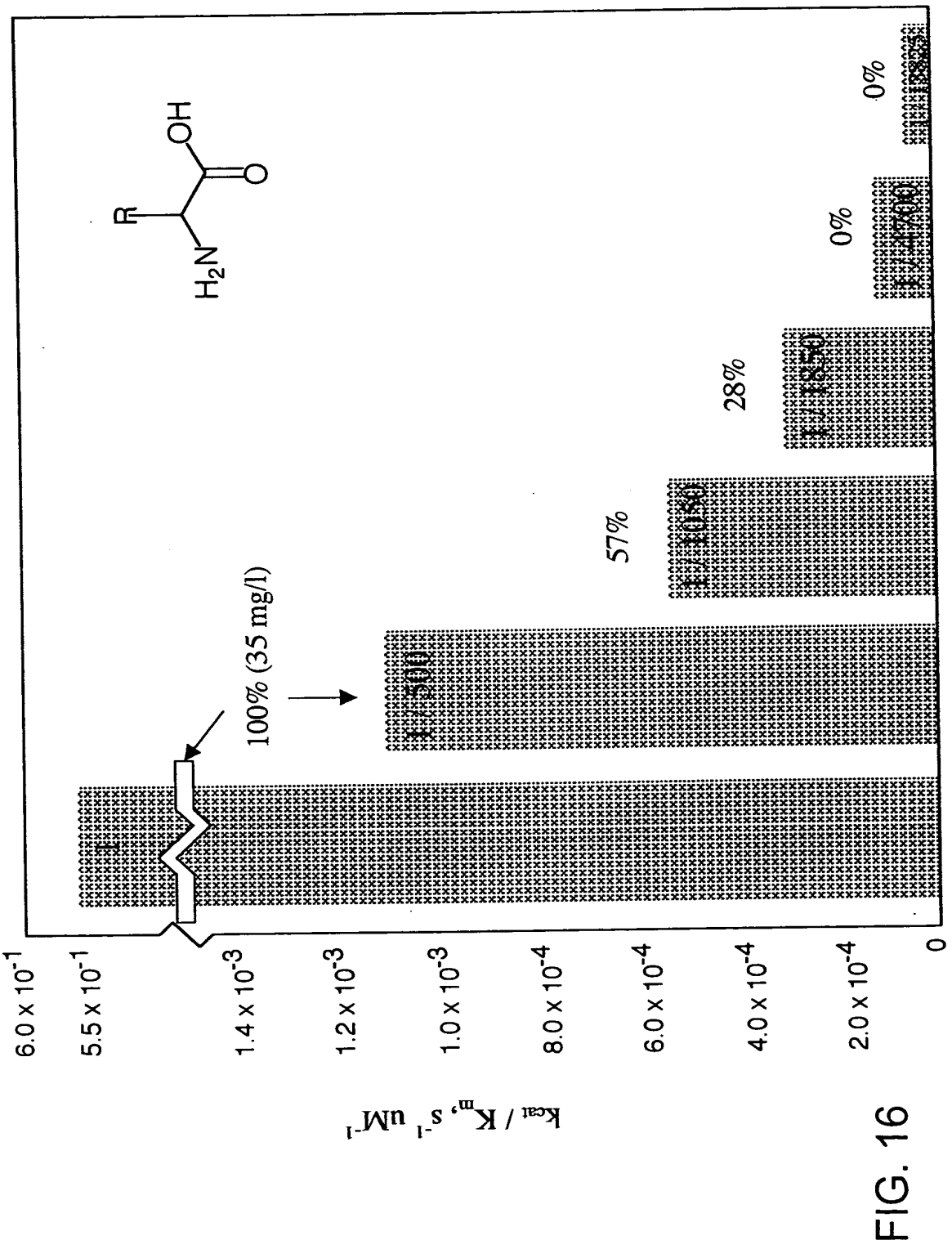

k_{cat} / K_m, s^{-1} uM^{-1}
6.0 x 10^{-1}
5.5 x 10^{-1}
1.4 x 10^{-3}
1.2 x 10^{-3}
1.0 x 10^{-3}
8.0 x 10^{-4}
6.0 x 10^{-4}
4.0 x 10^{-4}
2.0 x 10^{-4}
0
100% (35 mg/l)
57%
28%
0%
0%
R
OH
O
H_2N

FIG. 16

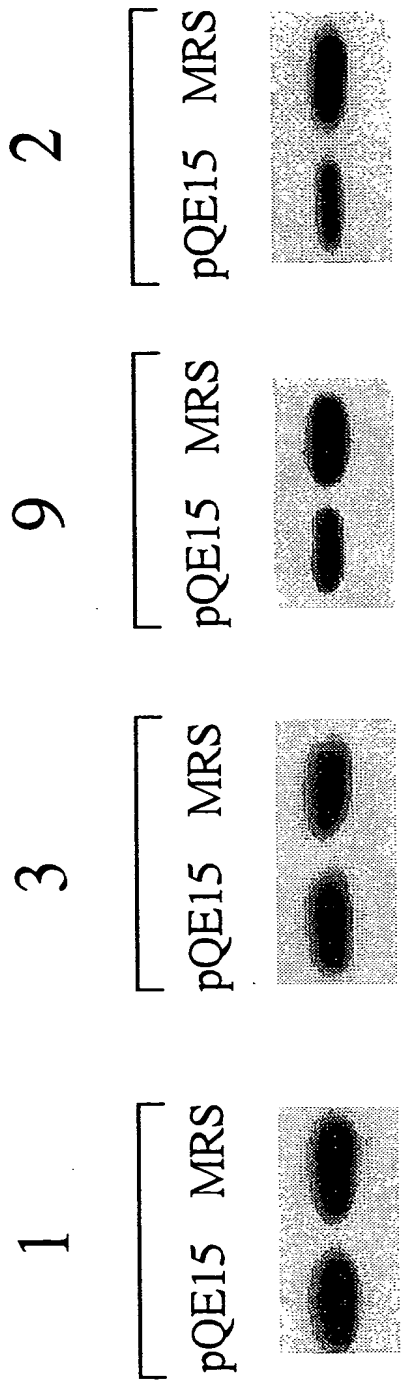


FIG. 17

a) $$AARS + aa + ATP \rightleftharpoons [AARS{:}aa{\sim}AMP] + PP_i$$

b) $$[AARS{:}aa{\sim}AMP] + tRNA^{aa} \longrightarrow aa{\sim}tRNA^{aa} + AMP + AARS$$

FIG. 18

```
CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA 60
ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA 120
GGATCGCATC ACCATCACCA TCACGGATCC GGCATCATGG TTCGACCATT GAACTCGATC 180
GTCGCCGTGT CCCAAAATAT GGGGATTGGC AAGAACGGAG ACCTACCCTG GCCTCCGCTC 240
AGGAACGAGT TCAAGTACTT CCAAAGAATG ACCACAACCT CTTCAGTGGA AGGTAAACAG 300
AATCTGGTGA TTATGGGTAG GAAAACCTGG TTCTCCATTC CTGAGAAGAA TCGACCTTTA 360
AAGGACAGAA TTAATATAGT TCTCAGTAGA GAACTCAAAG AACCACCACG AGGAGCTCAT 420
TTTCTTGCCA AAAGTTTGGA TGATGCCTTA AGACTTATTG AACAACCGGA ATTGGCAAGT 480
AAAGTAGACA TGGTTTGGAT AGTCGGAGGC AGTTCTGTTT ACCAGGAAGC CATGAATCAA 540
CCAGGCCACC TTAGACTCTT TGTGACAAGG ATCATGCAGG AATTTGAAAG TGACACGTTT 600
TTCCCAGAAA TTGATTTGGG GAAATATAAA CTTCTCCCAG AATACCCAGG CGTCCTCTCT 660
GAGGTCCAGG AGGAAAAAGG CATCAAGTAT AAGTTTGAAG TCTACGAGAA GAAAGGTTGG 720
AAGATCTTAA GCTTAATTAG CTGAGCTTGG ACTCCTGTTG ATAGATCCAG TAATGACCTC 780
AGAACTCCAT CTGGATTTGT TCAGAACGCT CGGTTGCCGC CGGGCGTTTT TTATTGGTGA 840
GAATCCAAGC TAGCTCTAGA GACGTCCGGC CGGAGCTCCA CCGCGGTGGC GGCCGCTCTA 900
GAGTCACTTA CTTAACATTT TCCCATTTGG TACTATCTAA CCCCTTTTCA CTATTAAGAA 960
GTAATGCCTA CTATGACTCA AGTCGCGAAG AAAATTCTGG TGACGTGCGC ACTGCCGTAC 1020
GCTAACGGCT CAATCCACCT CGGCCATATG CTGGAGCACA TCCAGGCTGA TGTCTGGGTC 1080
CGTTACCAGC GAATGCGCGG CCACGAGGTC AACTTCATCT GCGCCGACGA TGCCCACGGT 1140
ACACCGATCA TGCTGAAAGC TCAGCAGCTT GGTATCACCC CGGAGCAGAT GATTGGCGAA 1200
ATGAGTCAGG AGCATCAGAC TGATTTCGCA GGCTTTAACA TCAGCTATGA CAACTATCAC 1260
TCGACGCACA GCGAAGAGAA CCGCCAGTTG TCAGAACTTA TCTACTCTCG CCTGAAAGAA 1320
AACGGTTTTA TTAAAAACCG CACCATCTCT CAGCTGTACG ATCCGGAAAA AGGCATGTTC 1380
CTGCCGGACC GTTTTGTGAA AGGCACCTGC CCGAAATGTA AATCCCCGGA TCAATACGGC 1440
GATAACTGCG AAGTCTGCGG CGCGACCTAC AGCCCGACTG AACTGATCGA GCCGAAATCG 1500
GTGGTTTCTG GCGCTACGCC GGTAATGCGT GATTCTGAAC ACTTCTTCTT TGATCTGCCC 1560
```

FIG. 19-1

```
TCTTTCAGCG AAATGTTGCA GGCATGGACC CGCAGCGGTG CGTTGCAGGA GCAGGTGGCA 1620
AATAAAATGC AGGAGTGGTT TGAATCTGGC CTGCAACAGT GGGATATCTC CCGCGACGCC 1680
CCTTACTTCG GTTTTGAAAT TCCGAACGCG CCGGGCAAAT ATTTCTACGT CTGGCTGGAC 1740
GCACCGATTG GCTACATGGG TTCTTTCAAG AATCTGTGCG ACAAGCGCGG CGACAGCGTA 1800
AGCTTCGATG AATACTGGAA GAAAGACTCC ACCGCCGAGC TGTACCACTT CATCGGTAAA 1860
GATATTGTTT ACTTCCACAG CCTGTTCTGG CCTGCCATGC TGGAAGGCAG CAACTTCCGC 1920
AAGCCGTCCA ACCTGTTTGT TCATGGCTAT GTGACGGTGA ACGGCGCAAA GATGTCCAAG 1980
TCTCGCGGCA CCTTTATTAA AGCCAGCACC TGGCTGAATC ATTTTGACGC AGACAGCCTG 2040
CGTTACTACT ACACTGCGAA ACTCTCTTCG CGCATTGATG ATATCGATCT CAACCTGGAA 2100
GATTTCGTTC AGCGTGTGAA TGCCGATATC GTTAACAAAG TGGTTAACCT GGCCTCCCGT 2160
AATGCGGGCT TTATCAACAA GCGTTTTGAC GGCGTGCTGG CAAGCGAACT GGCTGACCCG 2220
CAGTTGTACA AAACCTTCAC TGATGCCGCT GAAGTGATTG GTGAAGCGTG GGAAAGCCGT 2280
GAATTTGGTA AAGCCGTGCG CGAAATCATG GCGCTGGCTG ATCTGGCTAA CCGCTATGTC 2340
GATGAACAGG CTCCGTGGGT GGTGGCGAAA CAGGAAGGCC GCGATGCCGA CCTGCAGGCA 2400
ATTTGCTCAA TGGGCATCAA CCTGTTCCGC GTGCTGATGA CTTACCTGAA GCCGGTACTG 2460
CCGAAACTGA CCGAGCGTGC AGAAGCATTC CTCAATACGG AACTGACCTG GGATGGTATC 2520
CAGCAACCGC TGCTGGGCCA CAAAGTGAAT CCGTTCAAGG CGCTGTATAA CCGCATCGAT 2580
ATGAGGCAGG TTGAAGCACT GGTGGAAGCC TCTAAATGAG AAGTAAAAGC CGCTGCCGCG 2640
CCGGTAACTG GCCCGCTGGC AGATGATCCG ATTCAGGAAA CCATCACCTT TGACGACTTC 2700
GCTAAAGTTG ACCTGCGCGT GGCGCTGATT GAAAACGCAG AGTTTGTTGA AGGTTCTGAC 2760
AAACTGCTGC GCCTGACGCT GGATCTCGGC GGTGAAAAAC GCAATGTCTT CTCCGGTATT 2820
CGTTCTGCTT ACCCGGATCC GCAGGCACTG ATTGGTCGTC ACACCATTAT GGTGGCTAAC 2880
CTGGCACCAC GTAAAATGCG CTTCGGTATC TCTGAAGGCA TGGTGATGGC TGCCGGTCCT 2940
GGCGGGAAAG ATATTTTCCT GCTAAGCCCG GATGCCGGTG CTAAACCGGG TCATCAGGTG 3000
AAATAATCCC CCTTCAAGGC GCTGCATCGA CAGCCTTTTG CTTTATAAAT TCCTAAAGTT 3060
GTTTTCTTGC GATTTTGTCT CTCTCTAACC CGCATAAATA CTGGTAGCAT CTGCATTCAA 3120
```

FIG. 19-2

```
CTGGATAAAA TTACAGGGAT GCAGAATGAG ACACTTTATC TATCAGGACG AAAAATCACA 3180
TAAATTCAGG GCAGTTGAGC AACAGGGAAA CGAGTTGCAT ATCAGTTGGG GAAAAGTTGG 3240
CACCAAAGGC AAAGCCAGAT AAAAAGTTTT TCAGATGCTG CGGCAGCGGC AAAAGCGGAG 3300
CCCGACCTCG AGGGGGGGCC CGGTACCCGG CCGGACGTCT CTAGAGCTAG CTTGGCGAGA 3360
TTTTCAGGAG CTAAGGAAGC TAAAATGGAG AAAAAAATCA CTGGATATAC CACCGTTGAT 3420
ATATCCCAAT GGCATCGTAA AGAACATTTT GAGGCATTTC AGTCAGTTGC TCAATGTACC 3480
TATAACCAGA CCGTTCAGCT GGATATTACG GCCTTTTTAA AGACCGTAAA GAAAAATAAG 3540
CACAAGTTTT ATCCGGCCTT TATTCACATT CTTGCCCGCC TGATGAATGC TCATCCGGAA 3600
TTTCGTATGG CAATGAAAGA CGGTGAGCTG GTGATATGGG ATAGTGTTCA CCCTTGTTAC 3660
ACCGTTTTCC ATGAGCAAAC TGAAACGTTT TCATCGCTCT GGAGTGAATA CCACGACGAT 3720
TTCCGGCAGT TTCTACACAT ATATTCGCAA GATGTGGCGT GTTACGGTGA AAACCTGGCC 3780
TATTTCCCTA AAGGGTTTAT TGAGAATATG TTTTTCGTCT CAGCCAATCC CTGGGTGAGT 3840
TTCACCAGTT TTGATTTAAA CGTGGCCAAT ATGGACAACT TCTTCGCCCC CGTTTTCACC 3900
ATGGGCAAAT ATTATACGCA AGGCGACAAG GTGCTGATGC CGCTGGCGAT TCAGGTTCAT 3960
CATGCCGTCT GTGATGGCTT CCATGTCGGC AGAATGCTTA ATGAATTACA ACAGTACTGC 4020
GATGAGTGGC AGGGCGGGGC GTAATTTTTT TAAGGCAGTT ATTGGTGCCC TTAAACGCCT 4080
GGGGTAATGA CTCTCTAGCT TGAGGCATCA AATAAAACGA AAGGCTCAGT CGAAAGACTG 4140
GGCCTTTCGT TTTATCTGTT GTTTGTCGGT GAACGCTCTC CTGAGTAGGA CAAATCCGCC 4200
GCTCTAGAGC TGCCTCGCGC GTTTCGGTGA TGACGGTGAA AACCTCTGAC ACATGCAGCT 4260
CCCGGAGACG GTCACAGCTT GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG 4320
CGCGTCAGCG GGTGTTGGCG GGTGTCGGGG CGCAGCCATG ACCCAGTCAC GTAGCGATAG 4380
CGGAGTGTAT ACTGGCTTAA CTATGCGGCA TCAGAGCAGA TTGTACTGAG AGTGCACCAT 4440
ATGCGGTGTG AAATACCGCA CAGATGCGTA AGGAGAAAAT ACCGCATCAG GCGCTCTTCC 4500
GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCTGTCGGC TGCGGCGAGC GGTATCAGCT 4560
CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG 4620
TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC 4680
```

FIG. 19-3

```
CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA 4740
AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT 4800
CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG 4860
GCGCTTTCTC AATGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG 4920
CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC CGGTAACTAT 4980
CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC 5040
AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC 5100
TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC 5160
GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT 5220
TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC 5280
TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG 5340
AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA 5400
ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA 5460
CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGCTG CCTGACTCCC CGTCGTGTAG 5520
ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGCGAGAC 5580
CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC 5640
AGAAGTGGTC CTGCAACTTT ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT 5700
AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGCCATTGC TACAGGCATC 5760
GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA ACGATCAAGG 5820
CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG TCCTCCGATC 5880
GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT 5940
TCTCTTACTG TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG 6000
TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT 6060
AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG 6120
CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA 6180
CCCAACTGAT CTTCAGCATC TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAAACAGGA 6240
```

FIG. 19-4

```
AGGCAAAATG CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC 6300
TTCCTTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA 6360
TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC CCGAAAAGTG 6420
CCACCTGACG TCTAAGAAAC CATTATTATC ATGACATTAA CCTATAAAAA TAGGCGTATC 6480
ACGAGGCCCT TTCGTCTTCA C                                          6501
```

FIG. 19-5

APPLICANT(S): Tirrell and Kiick
TITLE: OVEREXPRESSION OF AMINOACYL-tRNA SYNTHETASES FOR EFFICIENT PRODUCTION OF ENGINEERED PROTEINS CONTAINING AMINO ACID ANALOGUES
Application No.: Unassigned Filing Date: July 1, 2003
Docket No.: CIT1460-2

APPLICANT(S): Tirrell and Kiick
TITLE: OVEREXPRESSION OF AMINOACYL-tRNA SYNTHETASES FOR EFFICIENT PRODUCTION OF ENGINEERED PROTEINS CONTAINING AMINO ACID ANALOGUES
Application No.: Unassigned Filing Date: July 1, 2003
Docket No.: CIT1460-2


```
CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA 60
ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA 120
GGATCGCATC ACCATCACCA TCACGGATCC GGCATCATGG TTCGACCATT GAACTCGATC 180
GTCGCCGTGT CCCAAAATAT GGGGATTGGC AAGAACGGAG ACCTACCCTG GCCTCCGCTC 240
AGGAACGAGT TCAAGTACTT CCAAAGAATG ACCACAACCT CTTCAGTGGA AGGTAAACAG 300
AATCTGGTGA TTATGGGTAG GAAAACCTGG TTCTCCATTC CTGAGAAGAA TCGACCTTTA 360
AAGGACAGAA TTAATATAGT TCTCAGTAGA GAACTCAAAG AACCACCACG AGGAGCTCAT 420
TTTCTTGCCA AAAGTTTGGA TGATGCCTTA AGACTTATTG AACAACCGGA ATTGGCAAGT 480
AAAGTAGACA TGGTTTGGAT AGTCGGAGGC AGTTCTGTTT ACCAGGAAGC CATGAATCAA 540
CCAGGCCACC TTAGACTCTT TGTGACAAGG ATCATGCAGG AATTTGAAAG TGACACGTTT 600
TTCCCAGAAA TTGATTTGGG GAAATATAAA CTTCTCCCAG AATACCCAGG CGTCCTCTCT 660
GAGGTCCAGG AGGAAAAAGG CATCAAGTAT AAGTTTGAAG TCTACGAGAA GAAAGGTTGG 720
AAGATCTTAA GCTTAATTAG CTGAGCTTGG ACTCCTGTTG ATAGATCCAG TAATGACCTC 780
AGAACTCCAT CTGGATTTGT TCAGAACGCT CGGTTGCCGC CGGGCGTTTT TTATTGGTGA 840
GAATCCAAGC TAGCTCTAGA GACGTCCGGC CGGAGCTCCA CCGCGGTGGC GGCCGCTCTA 900
GAGTCACTTA CTTAACATTT TCCCATTTGG TACTATCTAA CCCCTTTTCA CTATTAAGAA 960
GTAATGCCTA CTATGACTCA AGTCGCGAAG AAAATTCTGG TGACGTGCGC ACTGCCGTAC 1020
GCTAACGGCT CAATCCACCT CGGCCATATG CTGGAGCACA TCCAGGCTGA TGTCTGGGTC 1080
CGTTACCAGC GAATGCGCGG CCACGAGGTC AACTTCATCT GCGCCGACGA TGCCCACGGT 1140
ACACCGATCA TGCTGAAAGC TCAGCAGCTT GGTATCACCC CGGAGCAGAT GATTGGCGAA 1200
ATGAGTCAGG AGCATCAGAC TGATTTCGCA GGCTTTAACA TCAGCTATGA CAACTATCAC 1260
TCGACGCACA GCGAAGAGAA CCGCCAGTTG TCAGAACTTA TCTACTCTCG CCTGAAAGAA 1320
AACGGTTTTA TTAAAAACCG CACCATCTCT CAGCTGTACG ATCCGGAAAA AGGCATGTTC 1380
CTGCCGGACC GTTTTGTGAA AGGCACCTGC CCGAAATGTA AATCCCCGGA TCAATACGGC 1440
GATAACTGCG AAGTCTGCGG CGCGACCTAC AGCCCGACTG AACTGATCGA GCCGAAATCG 1500
GTGGTTTCTG GCGCTACGCC GGTAATGCGT GATTCTGAAC ACTTCTTCTT TGATCTGCCC 1560
```

FIG. 20-1

```
TCTTTCAGCG AAATGTTGCA GGCATGGACC CGCAGCGGTG CGTTGCAGGA GCAGGTGGCA 1620
AATAAAATGC AGGAGTGGTT TGAATCTGGC CTGCAACAGT GGGATATCTC CCGCGACGCC 1680
CCTTACTTCG GTTTTGAAAT TCCGAACGCG CCGGGCAAAT ATTTCTACGT CTGGCTGGAC 1740
GCACCGATTG GCTACATGGG TTCTTTCAAG AATCTGTGCG ACAAGCGCGG CGACAGCGTA 1800
AGCTTCGATG AATACTGGAA GAAAGACTCC ACCGCCGAGC TGTACCACTT CATCGGTAAA 1860
GATATTGTTT ACTTCCACAG CCTGTTCTTC CCTGCCATGC TGGAAGGCAG CAACTTCCGC 1920
AAGCCGTCCA ACCTGTTTGT TCATGGCTAT GTGACGGTGA ACGGCGCAAA GATGTCCAAG 1980
TCTCGCGGCA CCTTTATTAA AGCCAGCACC TGGCTGAATC ATTTTGACGC AGACAGCCTG 2040
CGTTACTACT ACACTGCGAA ACTCTCTTCG CGCATTGATG ATATCGATCT CAACCTGGAA 2100
GATTTCGTTC AGCGTGTGAA TGCCGATATC GTTAACAAAG TGGTTAACCT GGCCTCCCGT 2160
AATGCGGGCT TTATCAACAA GCGTTTTGAC GGCGTGCTGG CAAGCGAACT GGCTGACCCG 2220
CAGTTGTACA AAACCTTCAC TGATGCCGCT GAAGTGATTG GTGAAGCGTG GGAAAGCCGT 2280
GAATTTGGTA AAGCCGTGCG CGAAATCATG GCGCTGGCTG ATCTGGCTAA CCGCTATGTC 2340
GATGAACAGG CTCCGTGGGT GGTGGCGAAA CAGGAAGGCC GCGATGCCGA CCTGCAGGCA 2400
ATTTGCTCAA TGGGCATCAA CCTGTTCCGC GTGCTGATGA CTTACCTGAA GCCGGTACTG 2460
CCGAAACTGA CCGAGCGTGC AGAAGCATTC CTCAATACGG AACTGACCTG GGATGGTATC 2520
CAGCAACCGC TGCTGGGCCA CAAAGTGAAT CCGTTCAAGG CGCTGTATAA CCGCATCGAT 2580
ATGAGGCAGG TTGAAGCACT GGTGGAAGCC TCTAAATGAG AAGTAAAAGC CGCTGCCGCG 2640
CCGGTAACTG GCCCGCTGGC AGATGATCCG ATTCAGGAAA CCATCACCTT TGACGACTTC 2700
GCTAAAGTTG ACCTGCGCGT GGCGCTGATT GAAAACGCAG AGTTTGTTGA AGGTTCTGAC 2760
AAACTGCTGC GCCTGACGCT GGATCTCGGC GGTGAAAAAC GCAATGTCTT CTCCGGTATT 2820
CGTTCTGCTT ACCCGGATCC GCAGGCACTG ATTGGTCGTC ACACCATTAT GGTGGCTAAC 2880
CTGGCACCAC GTAAAATGCG CTTCGGTATC TCTGAAGGCA TGGTGATGGC TGCCGGTCCT 2940
GGCGGGAAAG ATATTTTCCT GCTAAGCCCG GATGCCGGTG CTAAACCGGG TCATCAGGTG 3000
AAATAATCCC CCTTCAAGGC GCTGCATCGA CAGCCTTTTG CTTTATAAAT TCCTAAAGTT 3060
GTTTTCTTGC GATTTTGTCT CTCTCTAACC CGCATAAATA CTGGTAGCAT CTGCATTCAA 3120
```

FIG. 20-2

```
CTGGATAAAA TTACAGGGAT GCAGAATGAG ACACTTTATC TATCAGGACG AAAAATCACA 3180
TAAATTCAGG GCAGTTGAGC AACAGGGAAA CGAGTTGCAT ATCAGTTGGG GAAAAGTTGG 3240
CACCAAAGGC AAAGCCAGAT AAAAAGTTTT TCAGATGCTG CGGCAGCGGC AAAAGCGGAG 3300
CCCGACCTCG AGGGGGGGCC CGGTACCCGG CCGGACGTCT CTAGAGCTAG CTTGGCGAGA 3360
TTTTCAGGAG CTAAGGAAGC TAAAATGGAG AAAAAAATCA CTGGATATAC CACCGTTGAT 3420
ATATCCCAAT GGCATCGTAA AGAACATTTT GAGGCATTTC AGTCAGTTGC TCAATGTACC 3480
TATAACCAGA CCGTTCAGCT GGATATTACG GCCTTTTTAA AGACCGTAAA GAAAAATAAG 3540
CACAAGTTTT ATCCGGCCTT TATTCACATT CTTGCCCGCC TGATGAATGC TCATCCGGAA 3600
TTTCGTATGG CAATGAAAGA CGGTGAGCTG GTGATATGGG ATAGTGTTCA CCCTTGTTAC 3660
ACCGTTTTCC ATGAGCAAAC TGAAACGTTT TCATCGCTCT GGAGTGAATA CCACGACGAT 3720
TTCCGGCAGT TTCTACACAT ATATTCGCAA GATGTGGCGT GTTACGGTGA AAACCTGGCC 3780
TATTTCCCTA AAGGGTTTAT TGAGAATATG TTTTTCGTCT CAGCCAATCC CTGGGTGAGT 3840
TTCACCAGTT TTGATTTAAA CGTGGCCAAT ATGGACAACT TCTTCGCCCC CGTTTTCACC 3900
ATGGGCAAAT ATTATACGCA AGGCGACAAG GTGCTGATGC CGCTGGCGAT TCAGGTTCAT 3960
CATGCCGTCT GTGATGGCTT CCATGTCGGC AGAATGCTTA ATGAATTACA ACAGTACTGC 4020
GATGAGTGGC AGGGCGGGGC GTAATTTTTT TAAGGCAGTT ATTGGTGCCC TTAAACGCCT 4080
GGGGTAATGA CTCTCTAGCT TGAGGCATCA AATAAAACGA AAGGCTCAGT CGAAAGACTG 4140
GGCCTTTCGT TTTATCTGTT GTTTGTCGGT GAACGCTCTC CTGAGTAGGA CAAATCCGCC 4200
GCTCTAGAGC TGCCTCGCGC GTTTCGGTGA TGACGGTGAA AACCTCTGAC ACATGCAGCT 4260
CCCGGAGACG GTCACAGCTT GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG 4320
CGCGTCAGCG GGTGTTGGCG GGTGTCGGGG CGCAGCCATG ACCCAGTCAC GTAGCGATAG 4380
CGGAGTGTAT ACTGGCTTAA CTATGCGGCA TCAGAGCAGA TTGTACTGAG AGTGCACCAT 4440
ATGCGGTGTG AAATACCGCA CAGATGCGTA AGGAGAAAAT ACCGCATCAG GCGCTCTTCC 4500
GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCTGTCGGC TGCGGCGAGC GGTATCAGCT 4560
CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG 4620
TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC 4680
```

FIG. 20-3

```
CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA 4740
AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT 4800
CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG 4860
GCGCTTTCTC AATGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG 4920
CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC CGGTAACTAT 4980
CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC 5040
AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC 5100
TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC 5160
GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT 5220
TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC 5280
TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG 5340
AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA 5400
ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA 5460
CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGCTG CCTGACTCCC CGTCGTGTAG 5520
ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGCGAGAC 5580
CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC 5640
AGAAGTGGTC CTGCAACTTT ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT 5700
AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGCCATTGC TACAGGCATC 5760
GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA ACGATCAAGG 5820
CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG TCCTCCGATC 5880
GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT 5940
TCTCTTACTG TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG 6000
TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT 6060
AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG 6120
CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA 6180
CCCAACTGAT CTTCAGCATC TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAAACAGGA 6240
```

FIG. 20-4

```
AGGCAAAATG CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC 6300
TTCCTTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA 6360
TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC CCGAAAAGTG 6420
CCACCTGACG TCTAAGAAAC CATTATTATC ATGACATTAA CCTATAAAAA TAGGCGTATC 6480
ACGAGGCCCT TTCGTCTTCA C                                           6501
```

FIG. 20-5